RECENSEMENT DU CANADA DE 1996

IAC

Introduction au recensement et au codage automatisé

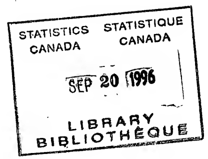


Codage automatisé

Statistique Canada Statistics Canada

Canada

IAC



Introduction au recensement et au codage automatisé

Rédigé par : Division des opérations du recensement
Secteur de la statistique sociale, des institutions et du travail



Nota : Pour alléger le texte, le genre masculin est employé la plupart du temps pour désigner à la fois les hommes et les femmes.

Table des matières

I. Introduction

Le présent manuel vous donne une vue d'ensemble du Recensement de la population de 1996 et du codage automatisé (CA).

II. Aperçu du recensement

A. Généralités

Le Recensement de la population est le plus important et le plus exhaustif des sondages menés au Canada. Il permet de recueillir des renseignements sur tous les habitants (hommes, femmes et enfants) du pays. La meilleure façon de le décrire est de parler d'un dénombrement officiel ou d'un inventaire national de tous les Canadiens.

Historiquement, le premier recensement effectué pour mesurer le progrès économique et social a été entrepris ici, au Canada, par Jean Talon. En 1666, Jean Talon a dénombré 3 215 personnes selon leur âge, leur sexe, leur état matrimonial et leur profession.

De nombreux recensements ont été effectués au Canada depuis 1666, mais le premier recensement décennal national date de 1871. En vertu de l'*Acte de l'Amérique du Nord britannique, 1867*, la tenue d'un recensement tous les 10 ans est une obligation constitutionnelle pour l'administration fédérale.

Avec l'ouverture des territoires de l'Ouest, de nouveaux colons se sont installés en grand nombre dans la région des Prairies. Afin de suivre cette croissance démographique fulgurante, des recensements quinquennaux (tous les cinq ans) ont été tenus par les administrations de ces provinces.

Inspirée par cet exemple, l'administration fédérale a jugé bon de recourir à un recensement quinquennal pour obtenir une mesure statistique au milieu de la période intercensitaire.

Depuis l'adoption, en 1971, de la *Loi sur la statistique*, l'administration publique canadienne effectue un recensement tous les cinq ans. La tenue d'un recensement quinquennal permet de refléter fidèlement l'évolution de la société. En outre, les utilisateurs de données peuvent procéder à des comparaisons dans le temps. Les questions du recensement ont été soigneusement conçues et mises à l'essai de sorte que la majeure partie des ménages puissent y répondre facilement. L'exactitude des résultats dépend de la bonne volonté des Canadiens à répondre sincèrement et avec précision aux questions du recensement. L'exactitude des résultats étant fonction de l'exhaustivitié des renseignements recueillis, le Parlement a stipulé que la participation au recensement était obligatoire. Parallèlement, il a incorporé à la *Loi sur la statistique* des dispositions relatives à la confidentialité des données qui interdisent de divulguer tout renseignement pouvant permettre d'identifier une personne.

Le dénombrement, comme la société, a beaucoup changé depuis l'époque de Talon. Pour répondre aux besoins accrus en matière de données socioéconomiques, il a fallu passer du simple comptage des personnes, effectué par Talon, à des méthodes plus élaborées comme l'autodénombrement, le dépouillement et le codage des renseignements fournis et le stockage des données sur support informatique.

La technique de l'échantillonnage, en vertu de laquelle quelques ménages seulement doivent remplir le questionnaire complet (2B), est utilisée aux fins du recensement, compte tenu de sa rentabilité et du fait qu'elle réduit le fardeau des répondants. Le questionnaire est distribué à un échantillon formé d'un ménage sur cinq, ou de 20 % de la population. Le reste de la population, soit 80 %, reçoit le questionnaire abrégé (2A). Ce dernier fournit les renseignements nécessaires à la prise de décisions de base. Il comprend sept questions ayant trait à l'âge, au sexe, à l'état matrimonial, aux unions libres et à la première langue apprise pendant l'enfance. Le questionnaire complet renferme les mêmes questions que le questionnaire abrégé, de même que des questions additionnelles portant notamment sur le revenu, le niveau de scolarité, l'origine ethnique, l'activité, la mobilité, la langue et le logement, soit 48 questions au total. Ces questions additionnelles permettent d'obtenir des renseignements importants mis à profit dans l'intérêt de tous les Canadiens.

On se sert des résultats du recensement pour prendre des décisions à l'échelle des quartiers, des collectivités, des provinces et du pays. Les administrations, les entreprises, les groupes communautaires, les prestataires de soins de santé, les chercheurs médicaux et les organismes d'un peu partout au Canada ont recours aux données du recensement pour régler diverses questions, depuis des politiques en matière de ressources humaines jusqu'à des programmes et des services locaux d'éducation, de formation, de promotion de la santé et de soutien à la collectivité. Par exemple, les données du recensement permettent de cerner les besoins futurs dans le domaine de l'emploi, ce qui facilite la planification des programmes d'éducation et de formation. En outre, grâce aux données du recensement, les décideurs peuvent déterminer les besoins eu égard aux routes, aux écoles, aux garderies, au transport public et aux programmes d'emploi pour les jeunes. De surcroît, le recensement vient étayer un certain nombre de lois et de statuts.

Le recensement constitue la principale source de statistiques au pays. Il sert de base au calcul des ratios, des indices, des taux, et autres. Pour chaque personne dénombrée, l'administration fédérale verse des paiements de transfert annuels aux provinces. Le non-dénombrement d'une personne entraîne chaque fois des pertes de centaines de dollars pour une province. Dans la même veine, les provinces et les territoires accordent aux administrations locales et municipales des subventions calculées d'après les chiffres de la population totale.

Les premières données du recensement sont diffusées six mois après que le travail sur le terrain ait pris fin. Pendant ces six mois, plus de 12,4 millions de questionnaires sont soumis aux étapes du traitement, du codage automatisé et de la compilation des données d'entrée, avant que les données résultantes puissent être analysées, imprimées et mises à la disposition de la population. Chaque étape doit être menée à terme et faire l'objet d'une certification avant que les données ne soient diffusées. Cette façon de procéder garantit que les données du recensement continuent d'atteindre les normes élevées que les utilisateurs en sont venus à exiger. Pour obtenir les données du recensement, il faut communiquer avec une des 52 bibliothèques de dépôt disséminées ça et là au pays ou avec une

des librairies qui vendent les publications gouvernementales. On peut se procurer des renseignements touchant le recensement en s'adressant aux services consultatifs des bureaux régionaux de Statistique Canada.

B. Limites du recensement

Pour réaliser un recensement dans un pays aussi vaste que le Canada, il est nécessaire de définir certaines régions géographiques de moindre importance afin de faciliter le dénombrement. Les régions géographiques de base utilisées aux fins du recensement sont les provinces (PROV), puis les circonscriptions électorales fédérales (CÉF) et finalement, le secteur de dénombrement (SD), qui est une unité encore plus petite. Un numéro de vérification (NV) est un chiffre de contrôle utilisé pour vérifier les dix numéros de PROV/CÉF/SD.
Les codes attribués aux provinces sont les suivants :

Terre-Neuve	= 10	Manitoba	= 46
Î.-P.-É.	= 11	Saskatchewan	= 47
Nouvelle-Écosse	= 12	Alberta	= 48
Nouveau-Brunswick	= 13	C.-B.	= 59
Québec	= 24	Yukon	= 60
Ontario	= 35	Territoires du Nord-Ouest	= 61

Une CÉF est une région géographique définie par une loi du Parlement, dont on se sert, dans le contexte du recensement, à des fins administratives. Chaque CÉF est en droit d'élire un député à la Chambre des communes. Les CÉF situées dans chacune des provinces se voient attribuer un code numérique à trois chiffres allant de 001 à 099.

Le SD correspond au secteur géographique dénombré par un recenseur. À des fins d'identification, chaque SD correspond à un code numérique précis. Les SD qui font partie des CÉF se voient également attribuer des codes numériques à trois chiffres allant de 001 à 999. Les logements privés (qui peuvent se composer d'un groupe familial avec ou sans autres personnes hors famille de recensement, de deux familles ou plus partageant le même logement, d'un groupe de personnes non apparentées ou d'une personne seule) sont numérotés de 001 à 799. Les codes attribuées aux logements collectifs (établissement commercial, institutionnel ou communautaire que le recenseur peut identifier comme tel grâce à une enseigne ou en s'adressant à la personne qui en a la charge, à un résident, à un voisin) sont numérotés de 901 à 999. En conséquence, chacune des régions géographiques dénombrées par un recenseur possédera un code d'identification à plusieurs chiffres unique. Ainsi, le numéro d'identification à 9 chiffres 10-006-036-7 (= PROV/CÉD/ SD/NV) correspond à :

10	Terre-Neuve
006	6e CÉF identifiée
036	36e SD identifié
7	numéro de vérification.

Pour l'ensemble du Canada, on compte environ 52 000 SD à chacun desquels on a attribué un numéro d'identification unique (PROV/CÉF/SD/NV).

C. Étapes du recensement

Le Recensement de la population est une entreprise très complexe étant donné qu'il s'écoule de sept à huit ans entre le moment où elle s'amorce et celui où elle s'achève et qu'elle mobilise de nombreux groupes d'employés de diverses divisions de Statistique Canada. La planification et la préparation initiales en vue du recensement débutent quatre années à l'avance afin de déterminer ce qui doit être fait et de planifier les ressources financières et humaines indispensables au projet.

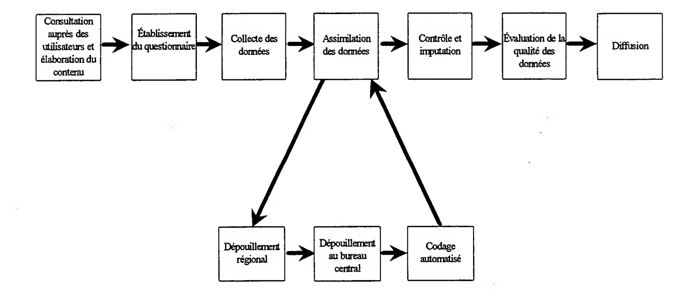
La planification d'un recensement commence par un exercice de planification stratégique qui consiste à remettre en question le contenu et le processus du recensement. Chaque aspect du processus du recensement peut être modifié. Les changements dépendent de l'évaluation du recensement précédent, des textes de loi et des politiques qui influent sur le recensement et des besoins et attentes de divers groupes publics qui s'intéressent aux données découlant du recensement.

Le processus du recensement comporte une série de systèmes et de procédures automatisés et manuels. Font partie de cette catégorie, l'élaboration du contenu et de la conception du questionnaire du recensement, la livraison des questionnaires du recensement aux répondants, le traitement des données recueillies, la mesure de la qualité des données et la diffusion des résultats.

Le Recensement du Canada de 1996 comporte sept grandes étapes :

- consultation auprès des utilisateurs et élaboration du contenu;
- établissement du questionnaire;
- collecte des données;
- assimilation des données;
- contrôle et imputation;
- évaluation de la qualité des données;
- diffusion.

Schéma de l'enchaînement des opérations du recensement de 1996



1. Consultation auprès des utilisateurs et élaboration du contenu

   Statistique Canada consulte les utilisateurs des données du recensement afin de déterminer le genre de renseignements dont ils ont le plus besoin. Il peut s'agir de renseignements sur le marché du travail ou le logement ou de données démographiques de base. Les consultations commencent d'ordinaire plusieurs années avant le jour du recensement.

2. Établissement du questionnaire

   Cette partie du cycle du recensement porte sur la détermination du contenu et la présentation finale des questionnaires du recensement, sur la mise en page et la préparation de la maquette, ainsi que sur l'impression, l'emballage et l'expédition des questionnaires aux bureaux régionaux.

3. Collecte des données

Le processus de collecte sur le terrain, mis en oeuvre par les recenseurs dans 52 000 secteurs de dénombrement (SD), comprend la livraison aux ménages canadiens et le retour par la poste d'environ 12 250 000 questionnaires (soit 9 800 000 formules 2A, 2 100 000 formules 2B, 1 750 formules 2C, 133 000 formules 2D et 200 000 formules 3). Ces questionnaires font ensuite l'objet d'un contrôle visant à déterminer s'ils sont dûment remplis par tous les Canadiens du pays. Cette opération occupe plus de 45 000 personnes.

4. Assimilation des données (AD)

Le processus de l'assimilation permet de convertir les renseignements consignés sur les questionnaires sous une forme lisible par machine. Pendant l'étape du dépouillement du cycle du recensement, les renseignements consignés sur les questionnaires font l'objet des trois opérations de production suivantes :

- Dépouillement régional (DR)

  Le DR consiste en la préparation manuelle et en la saisie sur clavier, sous une forme lisible par machine, des renseignements consignés sur tous les questionnaires du recensement. La préparation manuelle comporte les quatre opérations suivantes : réception et enregistrement, codage des variables économiques (industrie et profession), groupement par lots et étiquetage, enfin, expédition. Les données sont saisies au moyen d'un système de Revenu Canada (RC). Toutes les activités relatives au dépouillement régional sont mises en oeuvre par RC dans sept centres régionaux (St. John's, Jonquière, Shawinigan, Ottawa, Sudbury, Winnipeg et Surrey). Le DR occupe environ 1 900 personnes, soit 650 pour les activités de présaisie et 1 250 pour les activités de saisie.

- Dépouillement au bureau central (DBC)

  L'équipe du DBC est responsable de la réception, du contrôle et du stockage des questionnaires, des registres des visites (RV) et des cartouches de données. Le DBC consiste en une série d'opérations manuelles et automatisées conçues pour soumettre les données du recensement à des contrôles structurels. Il faut en outre dépouiller les questionnaires remplis par les Canadiens habitant à l'extérieur du pays (2C) et les personnes à bord de navires marchands et de navires de guerre. L'équipe du DBC transmet les données aux responsables du codage automatisé (CA) et du contrôle et imputation (C&I) et prépare un fichier de

données relatif aux chiffres provisoires et définitifs de population pour diffusion.

- Codage automatisé (CA)

  Le CA fait appel au traitement par lots et au traitement interactif pour apparier les réponses écrites figurant sur les questionnaires de recensement à des fichiers de référence ou structures de classification automatisés contenant une série de mots ou de syntagmes, en vue de leur attribuer un code numérique.

5. Contrôle et imputation (C&I)

   Au cours de cette étape, un système informatisé crée une base de données utilisable à partir des renseignements fournis par l'équipe de l'assimilation des données. Les préposés au contrôle et imputation vérifient les questions laissées sans réponse et les réponses invalides et illogiques, puis procèdent à l'imputation des données des répondants en vue de leur extraction et de leur publication ultérieures.

6. Évaluation de la qualité des données

   Ce processus a pour objet l'élaboration et la mise en oeuvre d'études sur la qualité des données du recensement, afin que les utilisateurs disposent de mesures (provenant des opérations de contrôle qualitatif) ou d'autres indicateurs de la fiabilité des résultats du recensement. Au nombre des études menées, mentionnons la contre-vérification des dossiers (CVD), la vérification des logements inoccupés (VLI) et l'étude de mesure de la couverture.

7. Diffusion

   L'étape finale consiste à extraire les données du recensement et à les présenter sous forme de tableaux devant figurer dans les publications et les totalisations personnalisées. Ces données sont diffusées sur divers supports, y compris les disques compacts ROM et les disquettes. L'équipe de la diffusion assume la responsabilité de l'élaboration de produits et de services qui répondent aux besoins des utilisateurs de données.

III. Codage automatisé (CA)

A. Processus

Le processus du codage automatisé est un amalgame d'étapes automatisées et manuelles, qui débutent par la réception des réponses écrites transmises par l'équipe du dépouillement au bureau central. Des enregistrements sont transmis renfermant des réponses alphanumériques, de même que d'autres données se rapportant à des personnes ou à des ménages particuliers qui peuvent être utiles pour faciliter le codage numérique de ces réponses.

Les réponses à 15 questions (variables), parmi les 48 qui figurent sur la formule 2B, seront envoyées au codage automatisé. À partir de ces 15 variables, on forme trois groupes principaux, soit :

- Lien avec la Personne 1
- Variables socioculturelles
  - Langue maternelle
  - Langue parlée à la maison
  - Langue non officielle
  - Lieu de naissance
  - Bande indienne/première nation
  - Origine ethnique
  - Lieu de résidence 1 an et 5 ans auparavant au Canada
  - Lieu de résidence 1 an et 5 ans auparavant à l'extérieur du Canada
  - Principal domaine d'études
  - Citoyenneté
  - Groupe de population
- Lieu de travail

Ces trois principaux groupes ont été formés en fonction des systèmes élaborés pour coder ces variables à l'étape du codage automatisé.

Un système central a été élaboré pour coder les variables socioculturelles. À l'origine, ce système a été conçu pour coder les variables socioculturelles dans le contexte du codage automatisé de 1991 et il a été perfectionné en vue de son utilisation à l'étape du codage automatisé de 1996.

Deux systèmes distincts tournant sur des ordinateurs personnels (FOXPRO) ont été élaborés pour coder à la fois la variable du lien avec la Personne 1 et celle du lieu de travail. Ces deux systèmes ont été conçus spécifiquement en vue du codage automatisé du recensement de 1996.

Les trois systèmes de codage ont pour objectif principal d'apparier les réponses écrites et les fichiers de référence déjà déterminés, ce qui permet d'attribuer un code numérique à la réponse écrite.

Ainsi, si un répondant donne l'«Algérie» comme lieu de naissance, le système parcourra le fichier de référence associé à la variable du lieu de naissance et y cherchera le mot «Algérie». Lorsqu'il trouve ce mot, le système lui attribue le code numérique correspondant, par exemple, 647.

Réponse écrite		Fichier de référence	
Algérie	non	Afghanistan	701
	non	Albanie	566
	oui	Algérie	647
		Samoa américaines	801

Si le système n'arrive pas à trouver le mot en question dans le fichier de référence (par exemple, si le répondant a incorrectement orthographié le mot ou si le mot est abrégé), il transmettra la réponse à l'étape suivante du codage automatisé qui est celle du codage interactif.

À l'étape du codage interactif, les codeurs examinent les réponses écrites que le système n'a pu coder et tentent de leur attribuer un code numérique en respectant les procédures déjà établies.

Une fois que toutes les réponses pour une variable ont été codées, par le système ou par les codeurs, on procède au contrôle qualitatif d'un échantillon et à l'évaluation dudit contrôle. Les codes ainsi obtenus sont alors stockés en vue d'être ultérieurement transmis à l'équipe du contrôle et imputation qui poursuivra le dépouillement.

Les trois systèmes élaborés possèdent de surcroît des caractéristiques distinctes. Vous trouverez, aux paragraphes suivants, un exposé de certaines des divergences entre ces systèmes.

- Système de codage du lien avec la Personne 1

  Les réponses sont codées pour deux types de ménages, selon qu'ils habitent dans un logement : privé ou collectif. Un logement privé peut se composer d'un groupe familial avec ou sans autres personnes hors famille de recensement, de deux ou plusieurs familles partageant un logement, d'un groupe de personnes non apparentées ou d'une personne seule. Un logement collectif s'entend d'un logement commercial, institutionnel ou communautaire que le recenseur peut identifier comme tel grâce à une enseigne ou en s'adressant à la personne qui en a la charge, à un résident ou à un voisin, qui est occupé par un ménage n'ayant pas de lieu habituel de résidence ailleurs au Canada.

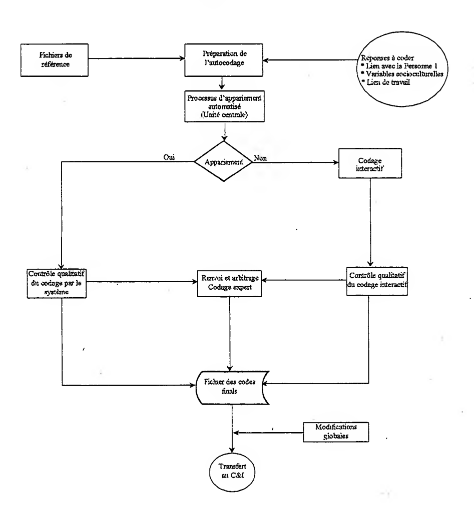
- Système de codage des variables socioculturelles

  Le système de codage des variables socioculturelles est fondé sur un logiciel d'ordinateur principal. L'appariement et le codage interactif s'effectuent tous deux sur les terminaux d'un ordinateur principal.

- Le système de codage du lieu de travail

  La principale différence entre le système de codage du lieu de travail et les deux autres systèmes élaborés en vue du codage automatisé est l'utilisation par le premier de cinq fichiers de référence plutôt qu'un pour procéder à l'appariement des réponses écrites. Le système explorera cinq fichiers de référence pour trouver un «lieu de travail» avant de décider s'il peut ou non apparier une réponse ou la transmettre aux codeurs. Ces derniers peuvent également consulter cinq fichiers de référence lorsqu'ils codent des réponses ayant trait au lieu de travail (LDT). Le code attribué à l'étape du codage du LDT correspond à un lieu géographique.

Codage automatisé de 1996
Enchaînement des opérations



B. Activités de codage

Les trois systèmes de codage interactif sont utilisés par des codeurs de niveaux différents. Les variables du lien avec la Personne 1 et les variables socioculturelles sont codées à deux niveaux, soit 1 et 2, par des codeurs généraux et experts. Le codage de la variable du lieu de travail s'effectue à trois échelons, par les codeurs généraux, des renvois et experts.

Les codeurs généraux (lien avec la Personne 1, variables socioculturelles et lieu de travail) codent manuellement les réponses mal orthographiées, les réponses abrégées, les cas particuliers comme celui-ci : «Indien, Autochtone, Même», les réponses multiples (cela se produit lorsqu'un répondant donne deux ou plusieurs réponses à une seule question) et les autres cas. Ils effectuent en outre le contrôle qualitatif des codes attribués par le système ou par les codeurs. Les réponses que les codeurs généraux sont incapables de coder feront l'objet d'un renvoi à l'échelon de codage supérieur.

Les codeurs des renvois (lieu de travail) codent les réponses renvoyées par les codeurs généraux. Pour ce faire, ils peuvent consulter des fichiers de référence additionnels auxquels seuls les codeurs des renvois et experts ont accès. Les codeurs des renvois agissent à titre d'arbitre dans les cas où le codeur général et celui du premier contrôle qualitatif ne s'entendent pas sur le code ou la réponse se rapportant à la variable du lieu de travail, c'est-à-dire qu'ils attribuent à celle-ci un nouveau code. Les codeurs des renvois attribuent un nouveau code à des «lots» complets de réponses codées par les codeurs généraux et rejetés au premier contrôle qualitatif. Finalement, les codeurs des renvois attribuent un code de lieu de travail aux réponses codées par d'autres codeurs des renvois et échantillonnées en vue d'un deuxième contrôle qualitatif.

Les codeurs experts (lien avec la Personne 1, variables socioculturelles et lieu de travail) effectuent les mêmes opérations de codage que les codeurs généraux et les codeurs des renvois. Ils se livrent en outre aux activités suivantes :

- ils codent les réponses renvoyées à un échelon de codage supérieur;
- ils codent les réponses dont le codage a été remis à plus tard. Ils trouvent le moyen de coder les réponses qui n'ont pu l'être lors d'une séance antérieure. Il s'agit des réponses devant faire l'objet d'une recherche plus poussée;
- ils codent à nouveau les réponses rejetées. Ils étudient et codent à nouveau les réponses des unités de travail rejetées qui ont été codées par des codeurs du niveau antérieur. Les réponses provenant d'une unité de travail rejetée qui doivent être à nouveau codées n'ont pas été sélectionnées pour faire partie de l'échantillon du contrôle qualitatif;

- ils effectuent le contrôle qualitatif. Ils examinent les réponses échantillonnées qui, à l'issue du codage du deuxième contrôle qualitatif, font partie de la catégorie des renvois en vue d'obtenir une majorité ou de celle des cas de triple divergence. Le code attribué par le codeur expert est comparé au code de production, au code QC1 et au code QC2. Une erreur est imputée au regard de chaque code qui ne correspond pas au code final attribué par l'expert;
- ils ont le droit d'attribuer le code final : «Incapable de coder». Le code attribué par un codeur expert est définitif et ne fait l'objet d'aucun contrôle qualitatif;
- ils s'occupent de la tenue et de la mise à jour des fichiers de référence.

IV. Codage automatisé et codage interactif – Contrôle qualitatif

Les réponses codées par le système et celles qui l'ont été par les codeurs font l'objet d'un contrôle qualitatif pour garantir un degré de qualité acceptable. Le contrôle qualitatif permet d'isoler les problèmes opérationnels liés aux codes attribués par le système ou les codeurs.

Le premier contrôle qualitatif effectué par un codeur consiste à attribuer un code à une réponse échantillonnée ayant déjà été codée par le système ou par un autre codeur. Le système procède ensuite à la comparaison des deux codes. S'ils sont identiques, le code est accepté et considéré comme définitif. En cas de divergence entre les deux codes (double divergence), la réponse est renvoyée au codeur du deuxième contrôle qualitatif qui attribue un code à la réponse échantillonnée. Le système compare alors ces trois codes. Si deux codes sont identiques, on considère qu'il s'agit du bon code et que ce choix est définitif. En outre, le code divergent est jugé erroné (règle de la majorité). Si tous les codes sont différents (triple divergence), la réponse est soumise à un codeur expert qui l'examinera (renvoi en vue d'obtenir une majorité).

Le codeur général ou le codeur des renvois ne sait pas si la réponse qui apparaît sur l'écran a déjà été codée ou non. Toutes les erreurs commises, que ce soit par le premier codeur général ou par le codeur général qui effectue le contrôle qualitatif, sont prises en note et font l'objet d'un suivi.

Le spécialiste du domaine responsable de chaque variable étudie les résultats et observe les taux d'erreurs.

V. Conclusion

Après avoir lu le manuel «Introduction au recensement et au codage automatisé», vous êtes prêt à passer à la prochaine section de la présente séance de formation, au cours de laquelle vous apprendrez à coder les réponses se rapportant à des variables particulières. Si des points demeurent obscurs, demandez des explications à votre surveillant.